

Rec'd PCT/PTO 24 SEP 2004  
PCT/IB 03 / 01040  
19.03.03



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

REC'D 31 MAR 2003

WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02290801.6

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

R C van Dijk

BEST AVAILABLE COPY



Anmeldung Nr:  
Application no.: 02290801.6  
Demande no:

Anmeldetag:  
Date of filing: 29.03.02  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.  
Groenewoudseweg 1  
5621 BA Eindhoven  
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se référer à la description.)

Video coding method and corresponding coded signal

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)  
revendiquée(s)  
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

H04N7/26

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

## "VIDEO CODING METHOD AND CORRESPONDING CODED SIGNAL"

### FIELD OF THE INVENTION

5           The present invention relates to the field of video compression and, for instance, to the video coding standards of the MPEG family (MPEG-1, MPEG-2, MPEG-4) and the ITU-H.26X family (H.261, H.263 and extensions, H.26L). More specifically, this invention concerns an encoding method applied to a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of  
10       said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels.

          The invention also relates to a corresponding encoding device, to a transmittable video signal consisting of a coded bitstream generated by such an encoding device, and to a  
15       device for receiving and decoding a video signal consisting of such a coded bitstream.

### BACKGROUND OF THE INVENTION

          In the first video coding standards (up to MPEG-2 and H.263), the video is assumed to be rectangular and to be described in terms of a luminance channel and two chrominance channels. With MPEG-4, other channels have been introduced : the alpha channel  
20       (also referred to as the "arbitrary shape channel" in MPEG-4 terminology), for describing the contours of the video objects, and, in a later version of MPEG-4, additional channels enabling the transmission of contents like depth, disparity or transparency. The depth, for instance, can be used for the applications where navigation in 3D is enabled. The disparity channel is used for the applications for which two views of the content are required, so that said content can be  
25       displayed on a device enabling stereoscopic viewing. The transparency channel is required for the contents composed of different objects which may be superimposed (a transparency channel for an object may be opaque – the object texture then overwrites the texture of the other objects - or half-transparent, the texture on the display then resulting from the blending of the texture of the objects).

          As defined in the MPEG-4-document w3056, "Information Technology – Coding of audio-visual objects – Part 2 : Visual", ISO/IEC/JTC1/SC29/WG11, Maui, USA, December 1999, part 6.2.3 Video Object Layer, the only way (in MPEG-4) to describe the additional channels like transparency or disparity or depth of a sequence is the use of the syntactic element  
30       "Video\_object\_layer\_shape\_extension". The syntax and the semantic provided by MPEG-4 in order to support the coding of additional channels via said element are given in pages 35-36  
35       and 110-112 of the document w3056 :

          (a) "video\_object\_layer\_verid" : this 4-bit code, defined in table 6-11, identifies the version number of the video object layer ;

(b) "video\_object\_layer\_shape" : this 2-bit code, defined in table 6-14, identifies the shape type of a video object layer ;

(c) "video\_object\_layer\_shape\_extension" : this 4-bit code, defined in table V2-1, identifies the number (up to 3) and type of auxiliary components that can be used (only a limited number of types and combinations are defined in said table, and more applications are possible by selection of the USER DEFINED type).

These syntax and semantic show that the support for the transmission of additional channels is only provided for objects having a shape. In case one wants to transmit the luminance and chrominance channels and one additional channel like the disparity of a rectangular object, it can indeed be explained how MPEG-4 is suboptimal in terms of coding efficiency. In MPEG-4, the description of a rectangular object (knowing that it is really rectangular since the code "video\_object\_layer\_shape" is then equal to 00) requires to transmit the size of the rectangle in terms of width and height. This description, which is given in the Video Object Layer syntax (see the five lines 25 to 30 of p.36 of the document), requires 31 bits. When one wants to transmit additional channels like the depth channel or the disparity channel of a rectangular object with the MPEG-4 syntax, there is no other means than to declare this object as non rectangular by setting the code "video\_object\_layer\_shape" to 11 (greyscale).

Once the object has been declared as being greyscale (although it is rectangular), the syntax forces to send bits describing the shape of the object, which is done at the macroblock level according to the syntax given in the document, p.52, § 6.2.6 Macroblock, lines 1 to 6 of the table, and p.56, § 6.2.6.1 MB Binary Shape Coding, lines 1 to 5 of the table. As indicated in p.128-129 of the document, bab\_type is a variable length code comprised between 1 and 7 bits and provided for indicating the coding mode used for the binary alpha block of 16 x 16 pixels, and the seven bab\_types are depicted in table 6-26. Such a description leads, for CIF pictures for instance, to a waste of bits at least 396 bits per frame (at least one bit per macroblock). For a 25 Hz CIF sequence, the overhead is estimated at 9,9 kbits/s.

## SUMMARY OF THE INVENTION

It is therefore an object of the invention to propose a video coding method allowing to avoid this waste of bits and therefore to improve the coding efficiency.

To this end, the invention relates to a method such as defined in the introductory part of the description and which is moreover characterized in that said syntax comprises a specific information indicating at a high description level in the bitstream the presence, or not, of the various channels that can be encountered to describe the content of the bitstream.

Preferably, said specific information consists of the following additional syntax elements :

video_object_layer_shape :	1 bit
number_of_video_object_layer_additional_channel_descriptions :	n bits

video\_object\_layer\_additional\_channels [i]

1 bit

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, the second one representing the number of additional channel syntax elements present in the coded bitstream in order to describe the content of said bitstream, and the third one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

In another embodiment of the invention, said specific information consists of the following additional syntax elements :

video\_object\_layer\_shape :

1 bit

number\_of\_video\_object\_layer\_additional\_channel\_presence :

n bits

video\_object\_layer\_additional\_channels [i]

1 bit

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, the second one representing the number of additional channels present in the coded bitstream, and the third one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

In a third embodiment, said specific information consists of the following additional syntax elements :

video\_object\_layer\_shape :

1 bit

video\_object\_layer\_additional\_channels [i]

1 bit,  $0 \leq i \leq 2^n-1$

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, the second one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

With anyone of these three solutions, the video\_object\_layer\_shape syntax element may be not provided in the bitstream.

The invention also relates to a device for encoding a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said device comprising means for structuring each scene of said sequence as a composition of video objects (VOs), means for coding the shape, the motion and the texture of each of said VOs, and means for multiplexing the coded elementary streams thus obtained into a single coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said device being further characterized in that it also comprises means for introducing into said coded bitstream a specific information indicating at a high description level in this coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

The invention also relates to a transmittable video signal consisting of a coded bitstream generated by an encoding method applied to a sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said coded bitstream, generated for coding all the video objects of said scenes, being constituted of encoded video data in which

each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said signal being further characterized in that said coded bitstream also comprises a specific information indicating at a high description level in this coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

The invention finally relates to a device for receiving and decoding a video signal consisting of a coded bitstream generated by an encoding method applied to a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said coded bitstream, generated for coding all the video objects of said scenes, being constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said coded bitstream moreover comprising a specific information indicating at a high description level in this coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

#### **DETAILED DESCRIPTION OF THE INVENTION**

To solve the problem of waste of bits explained above, it is proposed, according to the invention, to introduce into the coded bitstream an indication about the possible presence of additional channels. This indication consists of a specific information introduced, according to the invention, at a high description level at least equivalent to the Video Object Layer (VOL) MPEG-4 level.

This additional descriptive step is implemented for example as now indicated. The following syntactic elements are defined :

- (a) "video\_object\_layer\_shape" : 1 bit
- (b) "number\_of\_video\_object\_layer\_additional\_channel\_descriptions" : n bits
- (c) "video\_object\_layer\_additional\_channel [i]" : 1 bit

and the semantic meaning of these elements is :

(a) video\_object\_layer\_shape : this 1-bit flag indicates the presence of a shape (or contour) channel (if set to one, the contour channel is present and should be decoded, while no description of shape or contour is expected if it is not) ;

(b) number\_of\_video\_object\_layer\_additional\_channel\_descriptions : this n-bit unsigned integer represents the number of additional channel syntax elements present in the coded bitstream ;

(c) additional\_channel\_number : this integer takes values comprised between 0 and number\_of\_video\_object\_layer\_additional\_channel\_descriptions ;

(d) video\_object\_layer\_additional\_channel [additional\_channel\_number] :

this 1-bit flag identifies the presence or not of the channel addressed by the value [i] of additional\_channel\_number.

5 The correspondences between video\_object\_layer\_additional\_channel [additional\_channel\_number] and the semantic of the related channel are given in the following table, for values 1 to  $2^n$  of number\_of\_video\_object\_layer\_additional\_channel\_descriptions, called NAC in the table ( $n=4$  in the given example) :

Additional_channel_number	Semantic	No.of bits	NAC
0	video_object_layer_lum	1	1
1	video_object_layer_transparency	1	2
2	video_object_layer_disparity	1	3
3	video_object_layer_texture	1	4
4	video_object_layer_depth	1	5
5	user_defined	1	6
6	user_defined	1	7
7	user_defined	1	8
8	user_defined	1	9
9	user_defined	1	10
10	user_defined	1	11
11	user_defined	1	12
12	user_defined	1	13
13	user_defined	1	14
14	user_defined	1	15
.....	user_defined	1	.....

10 The proposition according to the invention leads therefore to a modified version of the syntax for Video\_object\_layer. In page 36 of the document w3056, the following syntactic elements are added (lines 15 and following) :

video_object_layer_shape	1	Uimsbf
if (video_object_layer_verid > 2) {		
number_of_video_object_layer_additional_channel_descriptions	n	Uimsbf
for (j=0 ; j< number_of_video_object_layer_additional_channel_descriptions, j++)		
video_object_layer_additional_channels[j]	1	uimsbf
}		

Examples of implementation (channel presence description + corresponding syntax) for various types of objects may be given, the syntax element which indicates the presence of chrominance channels being decoded only if the presence of a luminance channel has been indicated in the bitstream :

5	(a) a coloured 4:2:2 rectangular sequence :	
	video_object_layer_shape :	0
	number_of_video_object_layer_additional_channel_descriptions :	1
	video_object_layer_lum :	1
	video_object_layer_chrom :	1
10	(b) a black-and-white scene with an opaque object having a contour but no texture :	
	video_object_layer_shape :	1
	number_of_video_object_layer_additional_channel_descriptions :	0
	(c) a 4:2:2 black-and-white object having an opaque shape (or contour) :	
15	video_object_layer_shape :	1
	number_of_video_object_layer_additional_channel_descriptions :	1
	video_object_layer_lum :	1
	video_object_layer_chrom :	0
	(d) a coloured 4:2:2 rectangular object having a transparent alpha plane :	
20	video_object_layer_shape :	0
	number_of_video_object_layer_additional_channel_descriptions :	2
	video_object_layer_lum :	1
	video_object_layer_chrom :	1
	video_object_layer_transparency :	1
25	(e) a 4:2:2 rectangular object with its depth :	
	video_object_layer_shape :	0
	number_of_video_object_layer_additional_channel_descriptions :	5
	video_object_layer_lum :	1
	video_object_layer_chrom :	1
30	video_object_layer_transparency :	0
	video_object_layer_disparity :	0
	video_object_layer_texture :	0
	video_object_layer_depth :	1

35

The two following alternative syntaxes may also be proposed:

video_object_layer_shape	1	uimsbf
if (video_object_layer_verid > 2) {		
number_of_video_object_layer_additional_channel_presence	n	uimsbf
j = 0;		
k = 0;		
while (j < number_of_video_object_layer_additional_channel_descriptions)		
{		
j = j + video_object_layer_additional_channels[k];	1	uimsbf
k = k + 1;		
}		
}		

5

video_object_layer_shape	1	uimsbf
if (video_object_layer_verid > 2) {		
number_of_video_object_layer_additional_channel_descriptions = $2^n$ ;		
for (j=0 ; j<number_of_video_object_layer_additional_channel_descriptions, j++)		
video_object_layer_additional_channels[j]	1	uimsbf
}		

## CLAIMS :

1. An encoding method applied to a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said method being further characterized in that said syntax comprises a specific information indicating at a high description level in said coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

2. A method according to claim 1, in which said specific information consists of the following additional syntax elements :

video_object_layer_shape :	1 bit
number_of_video_object_layer_additional_channel_descriptions :	n bits
video_object_layer_additional_channels [i] :	1 bit

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, the second one representing the number of additional channel syntax elements present in the coded bitstream in order to describe the content of said bitstream, and the third one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

3. A method according to claim 1, in which said specific information consists of the following additional syntax elements :

video_object_layer_shape :	1 bit
number_of_video_object_layer_additional_channel_presence :	n bits
video_object_layer_additional_channels [i]	1 bit

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, the second one representing the number of additional channels present in the coded bitstream, and the third one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

4. A method according to claim 1, in which said specific information consists of the following additional syntax elements :

video_object_layer_shape :	1 bit
video_object_layer_additional_channels [i]	1 bit, $0 \leq i \leq 2^n-1$

the first element indicating the presence, or not, of a contour or shape channel that should then be decoded, and the second one identifying the presence, or not, of the channel addressed by the value [i], i taking a value between 0 and  $2^n-1$ .

5. A method according to anyone of claims 2 to 4, characterized in that the video\_object\_layer\_shape syntax element is not provided in the bitstream.

6. A device for encoding a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said device comprising means for

structuring each scene of said sequence as a composition of video objects (VOs), means for coding the shape, the motion and the texture of each of said VOs, and means for multiplexing the coded elementary streams thus obtained into a single coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said device being further characterized in that it also comprises means for introducing into said coded bitstream a specific information indicating at a high description level in said coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

7. ~~A transmittable video signal consisting of a coded bitstream generated by an encoding method applied to a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said coded bitstream, generated for coding all the video objects of said scenes, being constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said signal being further characterized in that said coded bitstream also comprises a specific information indicating at a high description level in said coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.~~

8. A device for receiving and decoding a video signal consisting of a coded bitstream generated by an encoding method applied to a video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs), said coded bitstream, generated for coding all the video objects of said scenes, being constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said coded bitstream moreover comprising a specific information indicating at a high description level in said coded bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream.

**Abstract**

The invention relates to an encoding method applied to a video sequence corresponding to successive scenes and generating a coded bitstream in which each data item is described by means of a bitstream syntax allowing, at the decoding side, to recognize and decode all the elements of the content of this coded bitstream. According to the invention, said syntax comprises a specific information indicating at a high description level in said bitstream the presence, or not, of various additional channels that can be encountered to describe the content of said bitstream. Several examples of specific information are given.

5

10

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**